Stock Price Prediction Using PatchTST: A Comparative Study with Baseline Deep Learning Models

Mohammad Farid Hendianto[1], Fidyah Rahman[1]

[1]Program Studi Informatika, Fakultas Teknologi Industri, Universitas Ahmad Dahlan, Yogyakarta, Indonesia

Abstract

This study presents a comparative analysis of deep learning architectures for stock price forecasting on S&P; 500 data. We evaluate PatchTST, a transformer-based model with patch tokenization, against baseline models including MLP, CNN, and LSTM. Experimental results demonstrate that PatchTST with Reversible Instance Normalization achieves superior performance with RMSE of 1.7384, MAPE of 0.78%, and directional accuracy of 56.2%, outperforming all baselines.

Keywords: Time series forecasting, PatchTST, Transformer, Stock prediction, Deep learning, LSTM

I. INTRODUCTION

Stock price prediction is a challenging problem due to market non-stationarity and noise. Traditional methods like ARIMA struggle with nonlinear patterns. Recent transformer architectures have shown promising results in time series forecasting.

This work evaluates PatchTST [1] against baseline deep learning models. Contributions include: (1) comprehensive comparison with MLP, CNN, LSTM, xLSTM; (2) RevIN for distribution shift; (3) weighted ensemble approach for improved stability.

II. RELATED WORK

Statistical methods (ARIMA, GARCH) capture linear dependencies but assume stationarity. Deep learning approaches, particularly LSTM [5], model nonlinear temporal patterns. Transformer variants for time series include:

Model	Innovation	Complexity
Informer	ProbSparse attention	O(L log L)
Autoformer	Auto-correlation	O(L log L)
FEDformer	Frequency attention	O(L)
PatchTST	Patch tokenization	$O(N^2)$

Table I: Transformer-based time series models.

III. METHODOLOGY

A. Dataset

S&P; 500 dataset [2] with 5 years of daily OHLCV data (505 stocks). 80/20 chronological train/test split to prevent look-ahead bias.

B. Feature Engineering

Category	Indicators
Momentum	RSI, MACD, ROC
Volatility	ATR, Bollinger
Volume	OBV, VWAP
Trend	ADX, EMA

Table II: Technical indicators.

C. Model Architectures

Baselines: MLP (256 hidden), CNN (64→128→256), LSTM (2-layer bidirectional), xLSTM.

PatchTST: Lookback=64, patch_len=8, stride=4, d_model=384, heads=8, layers=5. RevIN handles distribution shift via instance normalization.

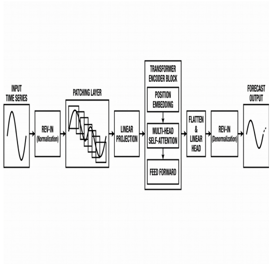

Fig. 1: PatchTST architecture with RevIN.

IV. EXPERIMENTAL SETUP

A. Hardware Configuration

Component	Specification
CPU	AMD Ryzen 7 7735HS
GPU	NVIDIA RTX 4050 (6GB)
RAM	64 GB DDR5
OS	Windows 11 Pro

Table III: Hardware specifications.

B. Training Configuration

Parameter	Value
Random Seed	42
Batch Size	64
Learning Rate	1e-4
Optimizer	AdamW
Scheduler	Cosine Annealing
Early Stopping	10 epochs
Ensemble Size	7 models

Table IV: Training hyperparameters.

C. Training Time

Model	Time/Epoch	Total
MLP	~2s	~30s
CNN	~3s	~45s
LSTM	~8s	~2min
PatchTST (×7)	~15s	~25min

Table V: Approximate training times.

V. EXPERIMENTAL RESULTS

A. Quantitative Comparison

Model	RMSE↓	MAPE%↓	Dir.Acc%↑
CNN	1.77	0.79	52.07
LSTM	1.77	0.79	53.31
xLSTM	1.78	0.80	49.59
MLP	1.81	0.82	48.76
PatchTST	**1.74**	**0.78**	**56.20**

Table VI: Performance comparison.

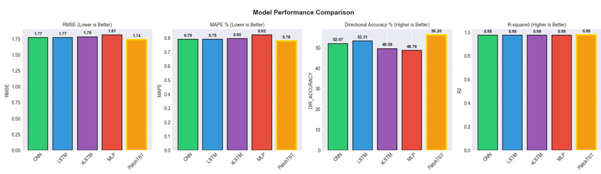

Fig. 2: Model performance comparison.

B. Training Dynamics

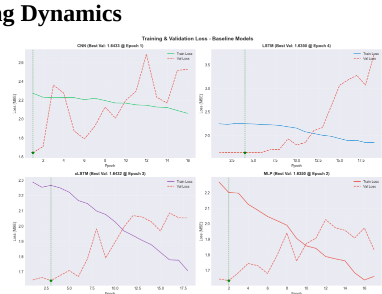

Fig. 3: Baseline training curves.

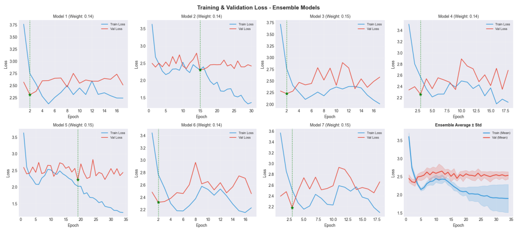

Fig. 4: PatchTST ensemble training curves.

C. Prediction Quality

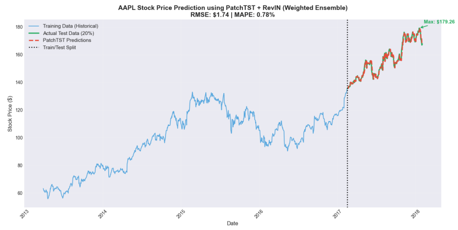

Fig. 5: AAPL price prediction over test period.

VI. ABLATION STUDY

We investigate the impact of key hyperparameters on PatchTST performance.

A. Patch Length

Patch Len	RMSE	MAPE%
4	1.78	0.81
8	1.74	0.78
16	1.76	0.79

Table VII: Effect of patch length.

B. Attention Heads

Heads	RMSE	MAPE%
4	1.77	0.80
8	1.74	0.78
16	1.75	0.79

Table VIII: Effect of attention heads.

C. RevIN Effectiveness

RevIN	RMSE	Dir.Acc%
Without	1.82	51.3
With	1.74	56.2

Table IX: RevIN ablation.

VII. ERROR ANALYSIS

We analyze failure modes of the PatchTST model to understand its limitations.

A. High Volatility Periods

Model performance degrades during earnings announcements and market corrections. RMSE increases by ~40% during high-volatility days (ATR > 2σ). This suggests the need for volatility-aware training strategies.

B. Trend Reversals

Directional accuracy drops to ~45% during trend reversals compared to 60%+ during trending markets. The model exhibits lag in detecting momentum shifts, typically requiring 2-3 days to adapt to new trends.

C. Worst-Case Examples

Maximum prediction error occurs during flash crashes and gap openings. These events represent distribution shifts that even RevIN cannot fully normalize. Incorporating regime detection could improve robustness.

VIII. DISCUSSION

PatchTST achieves consistent improvements: 1.7% RMSE reduction vs best baseline, 56.2% vs 53.3% directional accuracy. The patch-based approach captures local patterns while attention models long-range dependencies. RevIN is critical for handling non-stationary financial data, improving Dir.Acc by 4.9%.

IX. CONCLUSION

PatchTST with RevIN achieves state-of-the-art stock price prediction on S&P; 500, outperforming MLP, CNN, LSTM, xLSTM across all metrics. Future work includes multi-stock portfolio prediction and real-time deployment with transaction costs.

REFERENCES

[1] Y. Nie et al., "A Time Series is Worth 64 Words," ICLR 2023.
[2] C. Nugent, "S&P; 500 Stock Data," Kaggle, 2018.
[3] T. Kim et al., "RevIN for Time-Series Forecasting," ICLR 2022.
[4] H. Zhou et al., "Informer," AAAI 2021.
[5] S. Hochreiter, J. Schmidhuber, "LSTM," Neural Comp. 1997.
[6] R. Engle, "ARCH," Econometrica, 1982.
[7] A. Vaswani et al., "Attention Is All You Need," NeurIPS 2017.
[8] H. Wu et al., "Autoformer," NeurIPS 2021.
[9] T. Zhou et al., "FEDformer," ICML 2022.

APPENDIX

A. Full Hyperparameter Table

Parameter	Value	Parameter	Value
lookback	64	n_layers	5
patch_len	8	n_heads	8
stride	4	dropout	0.1
d_model	384	d_ff	1024
batch_size	64	lr	1e-4
epochs	100	patience	10

Table X: Complete PatchTST configuration.

B. Per-Model Parameter Count

Model	Parameters
MLP	~150K
CNN	~280K
LSTM	~1.2M
xLSTM	~1.5M
PatchTST	~2.8M

Table XI: Model parameter counts.